A COOPERATIVE UNSUPERVISED CONNECTIONIST MODEL TO IDENTIFY THE OPTIMAL CONDITIONS OF A PNEUMATIC DRILL

Emilio Corchado, Leticia Curiel, Pedro Bravo

Department of Civil Engineering, University of Burgos, Burgos, Spain.
E-mail: escorchardo@ubu.es

Abstract. A novel connectionist method to feature selection is proposed in this paper to identify the optimal conditions to perform drilling tasks. The aim is to extract information from complex high dimensional data sets. The model used is based on a family of cost functions which maximizes the likelihood of identifying a specific distribution in a data set. It employs lateral connections derived from the Rectified Gaussian Distribution to enforce a more sparse representation in each weight vector. The data investigated is obtained from the sensors allocated in a robot used to drill and build industrial warehouses. It is hoped that in classifying this data related with the strength, the water volume for refrigerating, speed and time of each sample, it will help in the search of the best conditions to perform the drilling of reinforce concrete slabs. This would produce a great saving for the company which owns the drilling robot.

1. Introduction

Feature extraction includes interesting tasks as space dimensionality reduction, feature construction, sparse representations and feature selection. All these techniques are commonly used as preprocessing in the field of machine learning. These areas have been tackled by researchers for many years and there has always been a great interest in feature extraction. The number of applications with very large input spaces which need space dimensionality reduction for efficiency and efficacy of the predictors is always increasing. These applications include very different areas as pattern recognition (e.g. handwriting recognition), speech processing, vision, text processing and recent ones as bioinformatics (DNA microarrays, etc.), and so on.

In this paper, we propose a method which is closely related to factor analysis. It is an unsupervised connectionist model based on the Negative Feed-

back Artificial Neural Network [1], which has been extended by the combination of two different techniques. Firstly by the selection of a cost function from a family of cost functions [2, 3] which identify different distributions. This method is called Maximum-Likelihood Hebbian learning [2, 3]. Secondly, competitive lateral connections [4] obtained from the Rectified Gaussian Distribution [5] were added to the Maximum-Likelihood method by Corchado et al. [2, 3] which enforced a greater sparsity in the weight vectors.

The Negative Feedback Artificial Neural Network has been linked to the statistical techniques of Principal Component Analysis [6] and Factor Analysis [7]. Lateral connections were introduced to the basic Maximum-likelihood network for the identification of different filters from video images [3, 4]. This paper reviews and proposes a method called Cooperative Maximum-Likelihood Hebbian Learning (CMLHL) [8, 9] to identify structure in a data set obtained from the sensors allocated in a robot used to drill and build industrial warehouses.

One of the biggest demands on present industrial field is the storage of goods in suitable places. So for this reason it is necessary to build up big warehouses for auto-carrier storage. Up to now, the drilling of slabs of reinforced concrete, which is necessary to place shelves on the mentioned warehouses, has been made manually by workers. This is a disadvantage due to the possible human errors which may produce a big economical loss. This is the main reason to design a robot that can carry out the mentioned operation raising the accuracy of the drilling tool. The robot must be capable of setting up the position of the bits and their diameters, to suck in, to filter and to recirculate the water for drilling, to test the tool conditions during automatic changing when it is required. In this way, it is easier to achieve a better assembly quality, decreasing the drilling execution time and achieving a less tool wear. All these factors imply an increase of the drilling instruments duration and the elimination of corrections. Finally this means the saving of a lot of money and time.

In order to identify the best conditions, we use an unsupervised connectionist method which can find sparse representations of the data which are more structured than those found by some other contemporary statistical techniques as Principal Component Analysis. It is also providing some global ordering.

This paper is organised as follows: from section 2 to 5 we introduce some statistical and connectionist techniques. Then we describe the problem and the data set. We finally show the results, conclusions and future work.

2. Principal Component Analysis

Principal component analysis (PCA) [10, 11] is a statistical method which aims to find the orthogonal basis which maximizes the variance of the projection of the data for a given dimensionality of basis. This usually means finding the direction which accounts for most of the data's variance; this becomes the first principal component. The next component is the direction from the remaining data which contains the most variance and is orthogonal to the previous basis vector.

This technique can be used as a dimension reduction technique which preserves as much information as possible in the remaining dimensions. If we consider only the largest eigenvalues corresponding to the principal components we can also find those components that contain most information, which may provide an insight into the structure of the data.

3. The Negative Feedback Neural Network

First we introduce the Negative Feedback Network [1], which is the basis of the Maximum-Likelihood model. Feedback is said to exist in a system whenever the output of an element in the system influences in part of the input applied to that particular element. It is used in this case to maintain the equilibrium on the weight vectors.

Consider an N-dimensional input vector, $\mathbf{X}$, and a M-dimensional output vector, $\mathbf{Y}$, with $\mathbf{W}_{ij}$ being the weight linking input $\boldsymbol{j}$ to output $\boldsymbol{i}$ and let η be the learning rate.

The initial situation is that there is no activation at all in the network. The input data is fed forward via weights from the input neurons (the $\mathbf{X}$-values) to the output neurons (the $\mathbf{Y}$-values) where a linear summation is performed to give the activation of the output neuron. This can be expressed as:

$$y_i = \sum_{j=1}^{N} W_{ij} x_j, \ \forall i \tag{1}$$

The activation is fed back through the same weights and subtracted from the inputs (where the inhibition takes place):

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \ \forall j \tag{2}$$

After that simple Hebbian learning is performed between input and outputs:

$$\Delta W_{ij} = \eta e_j y_i \tag{3}$$

The effect of the negative feedback is to stabilise the learning in the network. Because of that it is not necessary to normalise or clip the weights to get convergence to a stable solution.

This network is capable of finding the principal components of the input data in a manner that is equivalent to Oja's Subspace algorithm [6], and so the weights will not find the actual Principal Components but a basis of the Subspace spanned by these components.

Writing the algorithm in this way gives a model of the process which allows in general to envisage different models [2, 3, 4, 8] which would otherwise be impossible.

A method similar to PCA is Factor Analysis. Factor analysis (FA) is a multivariate technique whose goal is to represent a set of variables in terms of an underlying smaller set of variables called factors.

Its origins date back to the turn of the century to work by [12] concerned with understanding intelligence and the technique was developed for analysing the scores of individuals on a number of aptitude tests. Factor analysis has been developed mainly by psychologists and most applications

have been in the areas of psychology and the social sciences though others include medicine, geography and meteorology, for example. It attempts to explain the data set in terms of a smaller number of underlying factors. However Factor Analysis begins with a specific model and then attempts to explain the data by finding parameters which best fit this model to the data.

4. A Neural Exploratory Projection Pursuit Version

Exploratory Projection Pursuit (EPP) [13, 14] is a recent statistical method which is focused on solving the difficult problem of identifying structure in complex high dimensional data. It does this by projecting the data onto a low dimensional subspace in which the search for structure is done by eye. However not all projections will reveal the data's structure equally well. There is an index that measures how "interesting" a given projection is, and then represent the data in terms of projections that maximise that index.

Now "interesting" structure is usually defined with respect to the fact that most projections of high-dimensional data onto arbitrary lines through most multi-dimensional data give almost Gaussian distributions [15]. Therefore to identify "interesting" features in data, it is necessary to look for those directions onto which the data-projections are as far from the Gaussian as possible.

Corchado et al. [2, 3] presented a neural version of EPP in which the learning rule is given by:

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \approx y(p\,|\,\mathbf{e}\,|^{p-1}\ sign(\mathbf{e}))^{T} \tag{4}$$

Where T denotes the transpose of a vector. It is expected [2, 3] that for leptokurtotic residuals (more kurtotic than a Gaussian distribution), values of p<2 would be appropriate, while for platykurtotic residuals (less kurtotic than a Gaussian), values of p>2 would be appropriate.

Therefore the network operation is:

Feedforward: (Eq. 1)

Feedback: (Eq. 2)
Weight change: (Eq. 4)

By maximising the likelihood of the residual with respect to the actual distribution, there is a matching of learning rule to the pdf of the residual.

This method has been linked to the standard statistical method of Exploratory Projection Pursuit (EPP) [2, 3, 13, 14, 15]. EPP also gives a linear projection of a data set but chooses to project the data onto a set of basis vectors which best reveal the interesting structure in the data; interestingness is usually defined in terms of how far the distribution is from the Gaussian distribution. To identify interestingness, it is necessary to maximise the probability of the residuals under specific pdfs which are non-Gaussian.

5. Cooperative Maximum Likelihood Hebbian Learning

The Rectified Gaussian Distribution (RGD) [5] is a modification of the standard Gaussian distribution in which the variables are constrained to be non-negative, enabling the use of non-convex energy functions.
Lateral connections [5, 4] have been derived from the RGD and based on the cooperative distributions [5].

More formally, the standard Gaussian distribution is defined by:

$$p(\mathbf{y}) = Z^{-1} e^{-\beta E(\mathbf{y})}, \tag{5}$$

$$E(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T \mathbf{A}\mathbf{y} - \mathbf{b}^T \mathbf{y} \tag{6}$$

The quadratic energy function $\boldsymbol{E}(\mathbf{y})$ is defined by the vector $\mathbf{b}$ and the symmetric matrix $\mathbf{A}$. The parameter β=1/T is an inverse temperature. Lowering the temperature concentrates the distribution at the minimum of the energy function.
One advantage of this formalisation is that it allows to visualise regions of high or low probability in terms of energy and hence to view movement to low energy regions as movement to regions of high probability.

The quadratic energy function $\boldsymbol{E}(\mathbf{y})$ can have different types of curvature depending on the matrix A. Not all energy functions can be used in the Rectified Gaussian Distribution. The sorts of energy function that can be used are only those where the matrix A has the property:

$$\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y} > \mathbf{0} \text{ for all } \mathbf{y} : y_i > 0, i = 1...N \tag{7}$$

where N is the dimensionality of y. This condition is called co-positivity. This property blocks the directions in which the energy diverges to negative infinity.

The cooperative distribution in the case of N variables is defined by:

$$A_{ij} = \delta_{ij} + \frac{1}{N} - \frac{4}{N}\cos\left(\frac{2\pi}{N}(i-j)\right) \tag{8}$$

$$b_i = 1 \tag{9}$$

where δ_{ij} is the Kronecker delta and i and j represent the identifiers of output neuron.
To speed learning up, the matrix A can be simplified [7] to:

$$A_{ij} = \left(\delta_{ij} - \cos\left(2\pi(i-j)/N\right)\right) \tag{10}$$

The matrix A is used to modify the response to the data based on the relation between the distances between the outputs. The outputs are thought of as located on a ring ("wraparound").

The modes of the Rectified Gaussian are the minima of the energy function, subject to non-negativity constraints. The modes of the distribution characterize much of its behaviour at low temperature. It is possible to use what is probably the simplest algorithm, the projected gradient method, consisting of a gradient step followed by a rectification:

$$y_i(t+1) = \left[y_i(t) + \tau(b - \mathrm{Ay})\right]^+ \tag{11}$$

where the rectification $[\ ]^+$ is necessary to ensure that the **y**-values keep to the positive quadrant. If the step size τ is chosen correctly, this algorithm can provably be shown to converge to a stationary point of the energy function [16]. In practice, this stationary point is generally a local minimum.
The mode of the distribution can be approached by gradient descent on the derivative of the energy function with respect to **y**. This is:

$$\Delta \mathbf{y} \propto -\frac{\partial E}{\partial \mathbf{y}} = -(\mathrm{A}\mathbf{y} - \mathbf{b}) = \mathbf{b} - A\mathbf{y} \tag{12}$$

which is used as in Equation 11.
The final neural architecture will be shown to be an architecture which can find the independent factors of a data set but do so in a way which captures some type of global ordering in the data set.
The standard Maximum-Likelihood Network is used, but now with lateral connections (which acts after the feed forward but before the feedback). Then the final network is defined as follows:

Feedforward: (Eq.1)

Lateral Activation Passing:

$$y_i(t+1) = \left[y_i(t) + \tau(b - Ay)\right]^+ \tag{13}$$

Feedback: (Eq. 2)

Weight change: (Eq. 4)

Where the parameter τ represents the strength of the lateral connections.

6. Problem and Data Description

The purpose of the multidisciplinary work presented in this paper is the study of the best conditions for drilling reinforced concrete slabs, using a pneumatic drill with diamond bits, allocated in a robot.

In this work we have applied different unsupervised neural architectures in order to obtain the optimal conditions; the ones which cause least wear of the bits and provide the best result in less time. This will help the drilling company to save a lot of money.

The experiments are based on a data set obtained from a test done with reinforced concrete with steel bars. The data set has been collected under different circumstances, on the tool and also on the material to be drilled, in order to identify the optimal conditions in each case.

We have studied several variables and their response in a discrete range of values:

Variable (Units)	Range
□ Applied strength (N).	65, 80.5, 96, 11.5
□ Refrigerating volume water of the tool (l/min), which avoids its overheating and evacuate the waste.	2, 3, 4, 5
□ The speed of turn (r.p.m).	1000, 2000, 3000,4000
□ The drilling time (s).	80 different times.

Table 1. Variables, units and values used during the experiments. All these values have been chosen for been quite common for this drilling task.

The data set was obtained from the drilling of reinforced concrete test tubes with steel bars. The number of samples taken were relatively small, 80 samples. This is due to the high cost of the diamond bits.

7. Application to the Drilling Robot Data Set

The model presented above has been used to identify the optimal drilling conditions under different situations as:

- The bit faces only a concrete slab.
- The bit faces mainly the concrete slab and a small portion of a steal bar.

- The bit faces mainly a steal bar and the concrete slab.

Figure 1 shows results obtained using Cooperative Maximum Likelihood Hebbian Learning (CMLHL) and Principal Component Analysis. We can see that both methods have identified four different clusters.

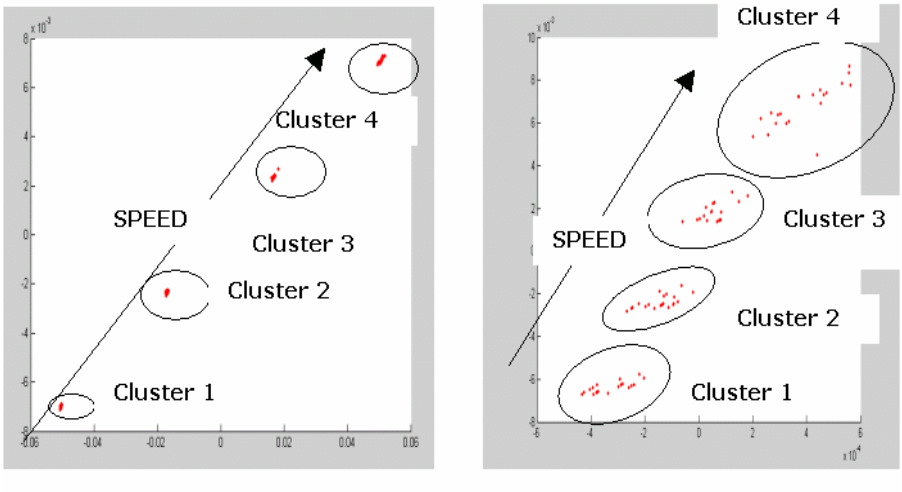

Fig. 1. Principal Component Analysis (left figure) and Cooperative Maximum Likelihood Hebbian Learning (right figure). CMLHL method identifies a projection which spreads the data out more than PCA.

Both methods have performed an initial classification attending to the speed factor. It is easy to affirm that CMLHL (Fig.1.- right figure) is providing a more sparse representation than PCA (Fig.1.-left figure). For a better understanding, we have studied the internal structure of the four clusters provided by CMLHL.

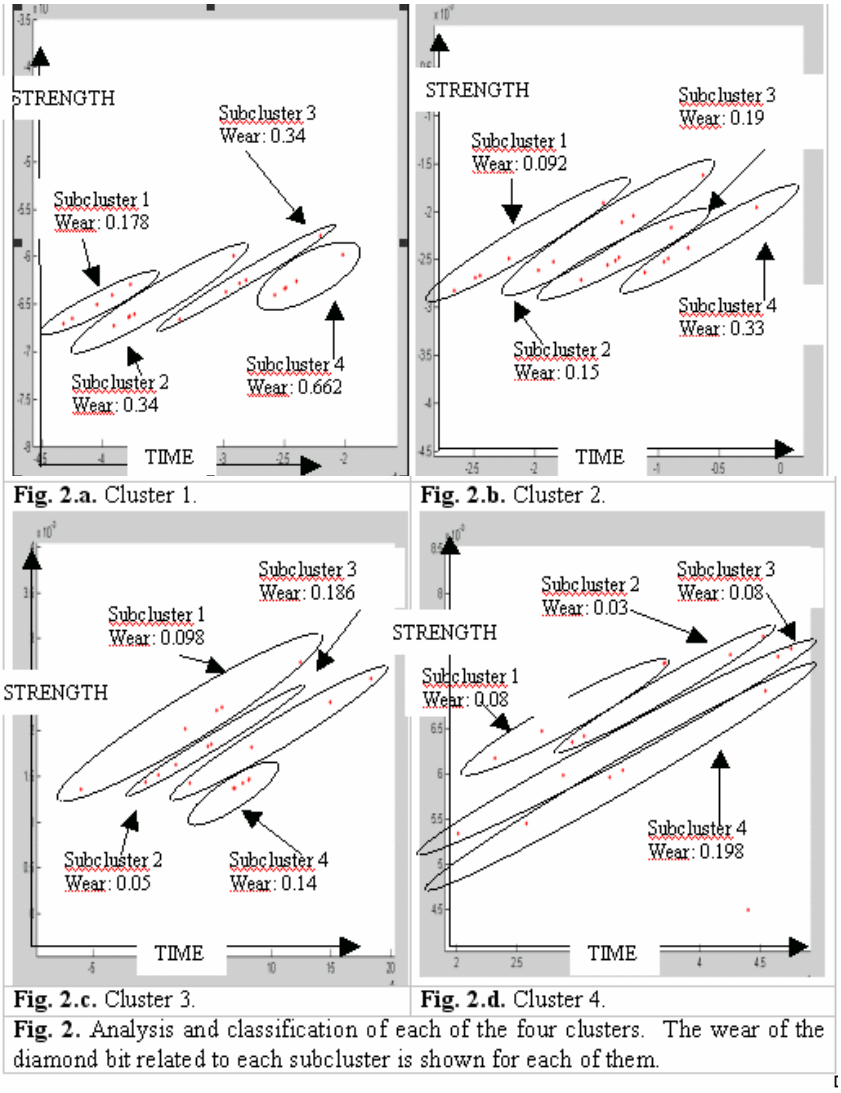

Fig. 2.a. Cluster 1.

Fig. 2.b. Cluster 2.

Fig. 2.c. Cluster 3.

Fig. 2.d. Cluster 4.

Fig. 2. Analysis and classification of each of the four clusters. The wear of the diamond bit related to each subcluster is shown for each of them.

After studying each subcluster we have noted a second classification. It is based on the strength applied and drilling time of each sample, which are decisive parameters. We have noticed that the ordinate axe (Fig 2.) in this case is related to the strength and that the coordinate axe (Fig 2.) is related to the applied time.

Once the results have been globally analyzed, we can state that the lower wear of the diamond bits take place in each cluster for subcluster one and two. The best conditions are related to subcluster 2 of cluster four (parameters: strength 80.5 N; 4000 r.p.m; time between 300 and 700 seconds and a medium volume) as we can see in Fig.2d. Of course, these results are relatives to the range of values used on the test, but in other hand quite common ones for this drilling task.

We would like to point out some interesting facts: The best conditions are not the ones related with the biggest values of strength. We have noticed that the use of a small refrigerating volume may produce an extreme warming of the bit and so for a bad elimination of the waste. In the opposite case, the use of an excessive amount of it may produce the breaking of the labs in a wrong way.

8. Conclusions

In this paper we have presented an ongoing multidisciplinary research in which we have showed an application of the CMLHL to a civil engineering problem and we have compared the method with a classical statistic method. We have identified the best conditions in a typical range of values and using a relatively quite small amount to samples.

Future work will be focus on the development of a product life management system which will record all the measures in a collaborative way. This will allow us the applications of artificial neural networks for different tasks related to the ones presented above and more important, to solve problems in real time.

9. References

[1] C. Fyfe A Neural Network for PCA and Beyond, Neural Processing Letters, 6:33-41. (1996)

[2] E. Corchado. D.MacDonald, and C. Fyfe, Maximun and Minimun Likelihood Hebbian Learning for Exploratory Projection Pursuit, Data Mining and Knowledge Discovery, Kluwer Academic Publishing. May 2004, vol. 8, no. 3, pp. 203-225. (2004)

[3] E. Corchado and C. Fyfe, Orientation Selection Using Maximum Likelihood Hebbian Learning, International Journal of Knowledge-Based Intelligent Engineering Systems Volume 7 Number 2, ISSN: 1327-2314. Brighton, United Kingdom. April (2003).

[4] E. Corchado, Y.Han, C. Fyfe, Structuring Global Responses of Local Filters Using Lateral Connections. J. Exp. Theor. Artif. Intell. 15(4): 473-487 (2003)

[5] H.S. Seung, N.D. Socci, and D. Lee, The Rectified Gaussian Distribution, Advances in Neural Information Processing Systems, 10. 350 (1998)

[6] E. Oja, Neural Networks, Principal Components and Subspaces, International Journal of Neural Systems, 1:61-68. (1989)

[7] D. Charles, Unsupervised Artificial Neural Networks for the Identification of Multiple Causes in Data. PhD thesis, University of Paisley (1999)

[8] E. Corchado, J.M Corchado, L. Sáiz and A. Lara, ., Constructing a Global and Integral Model of Business Management Using a CBR System. First International Conference on Cooperative Design, Visualization and Engineering CDVE2004, September 19-22, 2004, Palma de Mallorca. Spain. Page: 141- 147.

[9] E. Corchado, A. Herrero, B. Baruque, J.M Sáiz. Intrusion Detection System Based on a Cooperative Topology Preserving Method. ICANNGA 2005, International Conference on Adaptive and Natural Computing Algorithms. Coimbra. Portugal. 2005 (Accepted).

[10] Pearson, K. On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine, 2:559-572, 1901.

[11] Hotellind. H., Analysis of a Complex of Statistical Variables into Principal Components. Journal of Education Psychology, 24:417-444. 1933.

[12] Spearman, C., General intelligence, Objectively Determined and Measured, American Journal of Psychology. 1904.

[13] Hyvärinen, A. 2001. Complexity Pursuit: Separating interesting components from time series. Neural Computation, 13: 883-898.

[14] Friedman and J. Tukey. A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Transaction on Computers, (23): 881-890. (1974).

[15] Diaconis, P. and Freedman D. 1984. Asymptotics of Graphical Projections. The Annals of Statistics. 12(3): 793-815.

[16] Bertsekas D. P. Nonlinear Programming. Athena Scientific, Belmont, MA, (1995)